

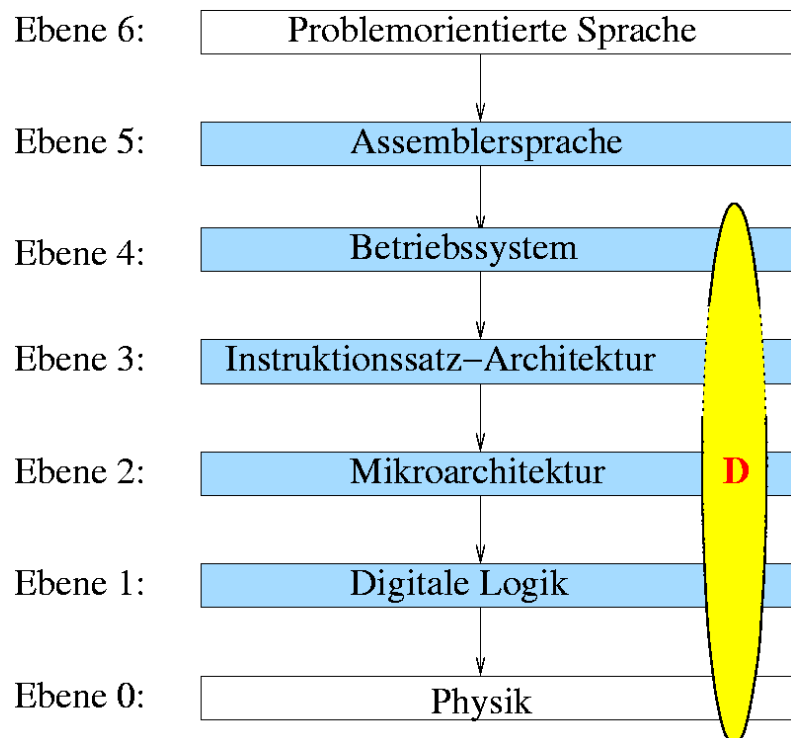
D Festplatten und Dateisysteme

1. Geschichte
2. Festplattenaufbau
3. Festplattenbusse: IDE /ATA und SCSI
4. CD-ROM und Floppy
5. RAID-Systeme
6. Dateisysteme

1

D Ein-/Ausgabekonzepte und Bussysteme

- Einordnung in das Schichtenmodell:

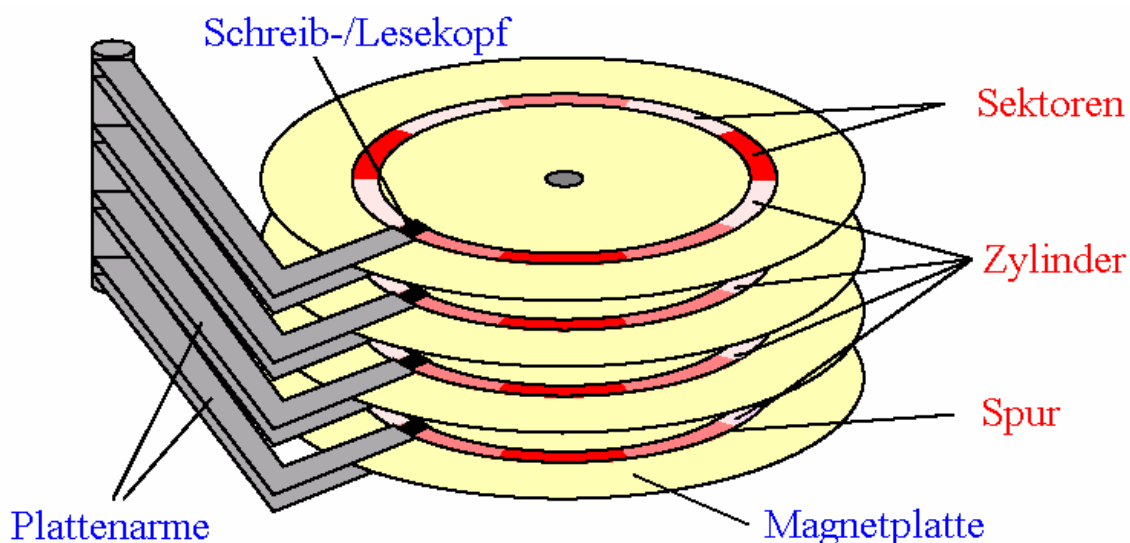


1 Geschichte

- kurzer Abriss der historischen Entwicklung:
 - 1956:** IBM stellt erste Festplatte vor (50 Scheiben mit je **24 Zoll**, insgesamt ca. 5 MByte Kapazität)
 - 1973:** 3340 Winchester Festplatte von IBM (Vorbild aller heutiger Platten)
 - 1979:** erste **5¼ Zoll** Festplatte von Seagate
 - 1981:** Einführung des SCSI-Peripheriebusses zum externen Anschluss von Festplatten
 - 1982:** Seagate stellt die ST506-Schnittstelle vor, aus der sich später IDE, ATA und ATAPI entwickelten
 - 1988:** erste **3½ Zoll** Festplatte von Conner Peripherals
 - 1992:** erste **2½ Zoll** Festplatte von PrairieTek, heute Standard für Laptops
 - 1996:** Seagate stellt Festplatten mit 10000 U/Min. vor
 - 1999:** IBM kündigt **1 Zoll** Festplatten (Microdrive) für portable Multimedia-Geräte an
- Speicherdichte stieg von ca. **300 Bit/cm²** im Jahre 1956 bis auf ca. **10 GBit/cm²** im Jahre 2003

2 Festplattenaufbau

- Aufbau einer Festplatte und Organisation der Daten (hier für eine Festplatte mit 4 Magnetplatten) :

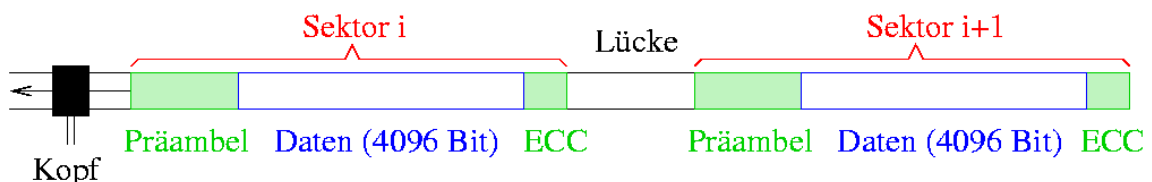


2 Festplattenaufbau (2)

- Aufbau einer Festplatte:
 - eine oder mehrere rotierende an einer gemeinsamen Achse (Spindel) befestigte **Magnetplatten** (unten und oben magnetisch beschichtet)
 - **Plattenarme** mit je zwei **Schreib-/Leseköpfen**, die
 - radial zwischen dem innersten und dem äußersten nutzbaren Radius bewegt werden können
 - mit sehr geringem Abstand ($< 0.1 \mu\text{m}$) über der Magnetplatte schweben
 - Lesen von Information erfolgt durch die **Induktion** eines Stromes im Schreib-/Lesekopf aufgrund der Rotation
 - Schreiben von Information erfolgt durch Magnetisierung der Oberfläche
 - Magnetplatten und Plattenarme mit Köpfen befinden sich im **hermetisch abgeschlossenen Gehäuse** (\Rightarrow Schutz vor Feuchtigkeit, Schmutz)
 - zwei kleine **Elektromotoren**:
 - Antriebsmotor zur Rotation der Platte (typisch 5400 bis 15000 U/Min.)
 - Schrittmotor zur Bewegung des Plattenarmes (als **Aktuator** bezeichnet)

2 Festplattenaufbau (3)

- **Organisation** der Daten auf einer Festplatte:
 - eine **Spur** stellt einen konzentrischen Kreis auf einer Magnetplatte dar; Breite einer Spur: $\leq 1 \mu\text{m}$, Spuren je Zentimeter: 1000 bis 10000
 - jede Spur ist in **Sektoren** (zumeist mit **512 Byte Nutzdaten**) unterteilt:



- jeder Sektor enthält noch 40 bis 100 Bytes für Verwaltungsinformationen, (**Präambel**) und Fehlererkennung/-korrektur (**ECC**)
- es können nur ganze Sektoren gelesen oder geschrieben werden!
- bei Einsatz mehrerer Magnetplatten bilden alle übereinander liegenden Spuren einen **Zylinder** (\Rightarrow mehr Daten ohne neue Kopfpositionierung!)
- jeder Sektor ist somit durch **Kopfnummer**, **Zylindernummer** und **Sektornummer** eindeutig identifiziert

2 Festplattenaufbau (4)

- heutige Festplatten verwenden das **Zone Bit Recording**:
 - da die äußere Spur wesentlich länger als die innere Spur ist, können bei gleicher Datendichte hier mehr Daten gespeichert werden
 - Einteilung der Plattenoberfläche in mehrere Zonen, so dass die **äußeren Spuren mehr Sektoren** erhalten **als die inneren Spuren**
 - höhere Datenrate im äußeren Bereich möglich, da hier je Umdrehung mehr Daten gelesen bzw. geschrieben werden können
(⇒ Zylinder werden daher von außen nach innen nummeriert und beschrieben)
 - innerste Spur wird oft als **Parkposition** verwendet (Autopark)
- Signale vom Lesekopf sind sehr klein:
 - richtige Erkennung einzelner Bits ist wegen **Rauschen** (Störungen durch Nachbarspur) nicht möglich
 - statt dessen werden Bitströme trickreich kodiert, so dass Fehler korrigiert werden können (**PRML**-Kodierung)

2 Festplattenaufbau (5)

- beim **Formatieren** einer Festplatte
 - werden Sektoren angelegt und mit Präambel beschrieben (enthält z.B. die Sektor-, Kopf- und Zylinder Nummer)
 - werden fehlerhafte Sektoren (*bad sectors*) erkannt
- einige **Leistungsmerkmale** von Festplatten:
 - **Kapazität**: Speicherkapazität für Nutzdaten (in GByte)
 - **Größe des Pufferspeichers** (typisch 512 kByte bis 16 MByte): eine oder mehrere Spuren werden zwischengespeichert und von hier nur teilweise oder in anderer Reihenfolge ausgelesen
 - **Latenzzeit**: mittlere Wartezeit bei Zugriff in der gleichen Spur (entspricht Zeit für eine halbe Umdrehung)
 - **Zugriffszeit** (typisch 3.5 bis 10 ms): Summe aus mittlerer Zeit für neue Kopfpositionierung und mittlerer Wartezeit
 - **Transferrate** (typisch 500 bis 800 Mbit/s): interne Transferrate zwischen Platte und Controller

2 Festplattenaufbau (6)

- Daten zweier aktueller 3½ Zoll Festplatten (Stand 2003):

	Seagate Barracuda 7200 ST 3160021	Seagate Cheetah 15K ST 373543LW
Kapazität	160 GByte	73.4 GByte
Magnetplatten / Köpfe	2 / 4	4 / 8
Umdrehungen / Minute	7200	15000
Latenzzeit	4.2 ms	2 ms
Zugriffszeit	8.5 ms	3.6 ms
Transferrate	683 MBit/s	891 MBit/s
Bytes je Sektor	512	512
Schnittstelle	UltraATA/100	Ultra320 SCSI
Pufferspeicher	2 MByte	8 MByte
Fehlerrate	1 aus 10 ¹⁴ Bits	1 aus 10 ¹⁵ Bits

(Quelle: www.seagate.com)

2 Festplattenaufbau (7)

- die Steuerelektronik einer Festplatte (Controller) besteht aus:
 - **Kodierer / Dekodierer:**
 - Transformation von analogen Kopfsignalen in serielle digitale Signale und umgekehrt gemäß der gewählten Kodierung (z.B. PRML)
 - **Datenformatierer:**
 - trennt Nutzdaten von Präambel und erkennt/korrigiert Fehler mittels ECC (beim Lesen) oder berechnet ECC (beim Schreiben)
 - wandelt serielle digitale Daten in parallele Daten um und umgekehrt
 - **Hostadapter:**
 - Schnittstelle der Festplatte zum Systembus bzw. Ein-/Ausgabebus mit Implementierung eines Übertragungsprotokolls zum Host
 - Annahme von Kommandos und ggf. überlappte Ausführung mehrerer Kommandos
 - Steuerung und ggf. Optimierung der Kopfbewegung
 - Pufferspeicher für Inhalt mehrerer Sektoren

3.1 Festplattenbusse: IDE/ATA

- **IDE** (*Integrated Drive Electronics*, 1986):
 - nur eine Bezeichnung von Compaq und Western Digital für die **Idee**, die aufwendige Ansteuerung von der Hauptplatine auf eine Steuerplatine in der Festplatte zu verlagern
 - auf Hauptplatine des Rechners ist nur einfache Dekodierung erforderlich
 - leistet eine einfache Anbindung von Laufwerken mit Steuerplatine an den ISA-Bus
- **ATA** (*Advanced Technology Attachment*, 1989):
 - Spezifikation eines ANSI **Standards** für IDE
 - legt physikalische und elektrische Eigenschaften der Schnittstelle fest
 - definiert Protokoll zwischen Host und Plattenlaufwerk
 - Unterstützung von DMA
 - gestattet maximale Transferrate von 4,2 MByte/s
 - ein Kanal mit Anschlussmöglichkeit von zwei Laufwerken

3.1 Festplattenbusse: IDE/ATA (2)

- Kommunikation zwischen Computer und Festplatte erfolgt über den IDE/ATA-Registersatz:
 - Steuerung über **8-Bit Kontrollregister** (Auswahl):
 - 1) **Fehlerregister**: enthält ggf. einen Fehlercode des letzten Zugriffs
 - 2) **Sektorzahl-Register**: Anzahl der zu lesenden/schreibenden Sektoren
 - 3) **Sektornummer-Register**: Auswahl eines Sektors im Zylinder
 - 4) **Zylindernummer-Register** (low Byte): Auswahl eines Zylinders
 - 5) **Zylindernummer-Register** (high Byte): dito
 - 6) **Laufwerk-Register**: Auswahl eines von zwei Laufwerken (Bit 4) und Kopfnummer (Bits 3 bis 0)
 - 7) **Kommando- und Statusregister**:
 - beim Schreibzugriff: Übermittlung eines von 50 möglichen Befehlen (z.B. *Identify Drive, Seek, Read Sector, Write Sector, Format Track*)
 - beim Lesezugriff: Statusregister, enthält Signale wie *drive busy, drive ready*
 - Datentransfer über ein **16-Bit Datenregister**

3.1 Festplattenbusse: IDE/ATA (3)

- zwei Verfahren zur Adressierung eines Sektors:
 - 1) **CHS-Modus** (*Cylinder/Head/Sector*):
 - 4-Bit Kopfnummer im Laufwerkregister: max. 8 Platten, 16 Köpfe
 - 16-Bit Zylinderadresse: max. 65536 Zylinder
 - 8-Bit Sektornummer: max. 256 Sektoren je Spur
 - bei einer Nutzdatengröße von 512 Byte je Sektor beträgt die maximale Festplattenkapazität $16 \times 65536 \times 256 \times 512 \text{ Byte} = 128 \text{ GByte}$
 - wegen beschränkter Sektor- und Zylinderanzahl und Zone Bit Recording werden heute CHS-Adressen einer logischen Geometrie angegeben und vom Controller in die physikalischen CHS-Adressen umgerechnet
 - 2) **LBA-Modus** (*Logical Block Address*):
 - Sektoren werden über logische 28-Bit Blocknummern angesprochen
 - ausgewählt, wenn Bit 6 im Laufwerk-Register auf 1 gesetzt wird
 - gleiche Einschränkung bei Plattenkapazität wie bei CHS-Modus

3.1 Festplattenbusse: IDE/ATA (4)

- IDE/ATA-Festplatten werden durch **40-polige Flachbandkabel** und 40-polige **Steckverbindungen** mit Hauptplatine verbunden
- je Kabel 3 Pfostenstecker (zum Anschluß von 2 Festplatten)
- Signale auf IDE/ATA-Bus (Auswahl):
 - Reset: Laufwerke zurücksetzen
 - D0-D15: 16 Datenleitungen
 - A0-A2: 3 Adressleitungen (zur Registerauswahl)
 - DMACK, DMARQ: DMA Leitungen
 - IRQ: Interrupt Request
 - I/O Read, I/O Write: Lese-/Schreibzugriff
 - IORDY: Bereitschaftsmeldung, ggf. Anforderung von Wartezyklen
 - PDIAG, DASP, SPSYNC: Festplattenspezifische Signale
- auch **80-polige Flachbandkabel** (mit 40 zusätzlichen Masseleitungen zur Abschirmung), nötig bei höheren Transferraten

3.1 Festplattenbusse: IDE/ATA (5)

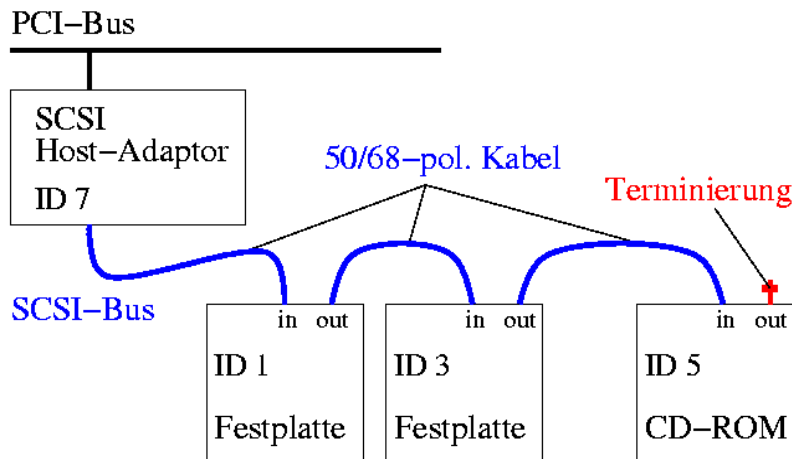
- Erweiterungen von **ATA** (Auswahl):
 - 1) **FastATA** oder **ATA-2** (1996):
 - Verdopplung der Transferrate auf 8.4 MByte/s
 - 2) **ATAPI** (*ATA Packet Interface*, auch in **ATA-4** enthalten):
 - Erweiterung des Standards zum Anschluss von CD-ROMs
 - 3) **UltraATA/66 (ATA-5)**, **UltraATA/100 (ATA-6)**, **UltraATA/133**
 - Transferraten von max. 66, 100 bzw, 133MByte/s
 - 4) **SerialATA**:
 - serielle Datenübertragung
 - einfachere Kabel, kleinere, preiswertere Steckverbinder
 - Transferrate bis zu 150 MByte/s
- **EIDE** (*Enhanced IDE*) ist kein Standard!
 - Oberbegriff von Western Digital für einige ATA-Erweiterungen; zusätzlich Definition von zwei Kanälen mit je zwei Festplatten

3.2 Festplattenbusse: SCSI

- **SCSI** (*Small Computer System Interface*) wurde 1979 von der Firma Shugart entwickelt (ursprünglich als SASI bezeichnet) und 1986 von ANSI als SCSI-1 standardisiert
(viele Anbieter haben sich zur SCSITA = *SCSI Trade Association* zusammengeschlossen ⇒ www.scsita.org)
- paralleles **Bussystem** zum Anschluss **diverser Peripheriegeräte** (vor allem Platten, CD-ROMs, Scanner und Bandgeräte):
 - es gibt **8-Bit** und **16-Bit** SCSI-Busse
 - Unterstützung von 8 Geräten (bei 8 Bit) bzw. 16 Geräten (bei 16 Bit) je Bus, die alle **gleichberechtigt** sind
 - Hostrechner (PC, Workstation) repräsentiert ein Gerät am SCSI-Bus; auch mehrere Rechner an einem SCSI-Bus prinzipiell möglich
 - direkter Datentransport zwischen den Geräten möglich
 - jedes SCSI-Gerät kann einen Bustransfer initiieren

3.2 Festplattenbusse: SCSI (2)

- Aufbau und Verkabelung eines SCSI-Busses:



- jedes SCSI-Gerät hat i.a. zwei gleichwertige Stecker zum Aufbau einer Verkettung
- **Terminierung**: das letzte Gerät am SCSI-Bus benötigt aus elektrischen Gründen einen Abschlusswiderstand

3.2 Festplattenbusse: SCSI (3)

- jedem SCSI-Gerät wird über einen Schalter (z.B. Jumper) eine **eindeutige Nummer** (SCSI ID) zugewiesen
(ein SCSI-Gerät kann zudem aus mehreren logischen Geräten bestehen, die durch LUNs = *Logical Unit Numbers* unterschieden werden, z.B. mehrere Partitionen einer Festplatte)
- **Busarbitrierung** auf SCSI-Bus:
 - möchte SCSI-Gerät mit ID i Initiator werden, so aktiviert es zunächst die i -te Datenleitung (sofern Bus frei ist)
 - wird innerhalb einer festen Zeitspanne ($2.4 \mu\text{s}$) keine andere Datenleitung aktiviert, so wird SCSI-Gerät mit ID i Initiator und **setzt** Signal **BSY** (bei einem Konflikt gewinnt SCSI Gerät mit höchster ID)
 - Bustransfer kann nicht unterbrochen werden; erst durch **Rücknahme** von **BSY** wird Bus explizit freigegeben

3.2 Festplattenbusse: SCSI (4)

- ein SCSI-Bustransfer erfolgt in mehreren Phasen:
 - 1) **Bus Free**: SCSI-Bus ist von keinem Gerät belegt
 - 2) **Arbitration**: Busarbitrierung, vgl. letzte Folie
 - 3) **Selection**: *Initiator i* ruft *Target j* durch Aktivierung der *j*-ten Datenleitung
 - 4) **Message Out**: *Initiator i* kann Mitteilung an *Target j* senden
 - 5) **Command**: *Initiator i* sendet ein Buskommando an *Target j*, das aus einem 1-Byte Befehlscode (z.B. Inquiry oder Read) und gerätespezifischen Parametern (z.B. Zylinder-, Kopf- und Sektornummer) besteht
 - 6) **Data In** oder **Data OUT**: eigentlicher Datentransfer
 - 7) **Status**: *Target* informiert *Initiator* über Status des Transfers
 - 8) **Message In**: *Target j* kann weitere Mitteilungen an *Initiator i* senden
- Geräte können **asynchronen** Bustransfer (Quittierung nach jedem Wort) oder **synchronen** Bustransfer (Quittierung nach vereinbartem Zeitintervall) vereinbaren

3.2 Festplattenbusse: SCSI (5)

- ein **8-Bit SCSI-Bus** benötigt einen **50-poligen** Stecker:
 - 8 Datenleitungen und 1 Leitung für Paritätsbit
 - 9 Steuerleitungen, z.B.:
 - **BSY**: SCSI-Bus belegt (*Busy*)
 - **SEL**: Anzeige der **Selection** Phase
 - **MSG**: Anzeige einer **Message** Phase
 - **ACK**: Bestätigung eines Transfers
 - **RST**: Rücksetzen *aller* SCSI-Geräte am Bus (*Reset*)
 - **ATN**: Initiator bittet um Aufmerksamkeit (*Attention*)
 - jede Signalleitung ist doppelt vorhanden:
 - Masseleitung zur Abschirmung für Betriebsart SE (*Single Ended*)
 - für differentielle Kodierung bei Betriebsart LVD (*Low Voltage Differential*)
 - +5V Spannungsleitung (nötig für Terminierung !)
- ein **16-Bit SCSI-Bus** benötigt **68-poligen** Stecker:
 - 8 weitere Datenleitungen
 - 1 weitere Leitung für Paritätsbit

3.2 Festplattenbusse: SCSI (6)

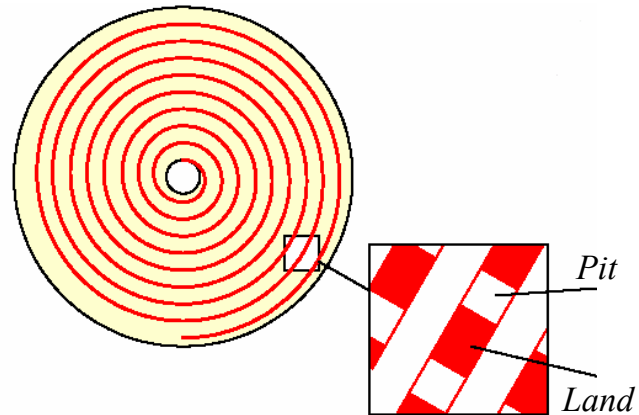
- verschiedene SCSI-Standards:
 - **SCSI-1** (1986): **5 MHz Bustakt**, 8-Bit Bus, 5 MByte/s, SE
 - **SCSI-2** (1990, Fast SCSI): **10 MHz Bustakt**, 8- oder 16-Bit Bus, SE
 - 10 MByte/s bei 8-Bit (*Fast narrow SCSI*)
 - 20 MByte/s bei 16-Bit (*Fast wide SCSI*)
 - erweiterter Befehlssatz z.B. zur Steuerung von CD-ROMs
 - **SCSI-3** (1996) ist eine Sammlung mehrerer Standards:
 - **Ultra**: **20 MHz Bustakt**, 20 MByte/s bei 8-Bit (*Ultra narrow SCSI*) und 40 MByte/s bei 16-Bit (*Ultra wide SCSI*), SE
 - **Ultra2**: **40 MHz Bustakt**, bis zu 80 MByte/s bei 16-Bit (*Ultra2 wide SCSI*), längere Leitungen (bis zu 25m) durch LVD
 - **Ultra160**: 40 MHz Bustakt, Datenübertragung bei **steigender** und **fallender** Taktflanke, LVD, max. 160 MByte/s bei 16-Bit
 - **Ultra320**: **80 MHz Bustakt**, LVD, bis zu 320 MByte/s
 - **Serial SCSI**: serielles Kabel (bis zu 10 km Länge); max. 400 MByte/s

3.3 Festplattenbusse: SCSI und IDE/ATA

- Vergleich von SCSI und IDE/ATA:
 - + SCSI erlaubt den Anschluss von Platten **beliebiger Kapazität**
 - + SCSI erlaubt zur Zeit (2003) **höhere Übertragungsraten** als IDE (sinnvoll jedoch i.a. nur beim Betrieb mehrerer Platten)
 - + SCSI erlaubt **8 bis 16 Geräte je Bus**, IDE nur zwei Geräte je Kanal
 - + SCSI erlaubt je Gerät eine **Warteschlange** für bis zu 256 Kommandos, die überlappt und ggf. in anderer Reihenfolge abgearbeitet werden können (*Command Queueing and Reordering*)
 - SCSI benötigt **zusätzlichen Host-Adaptor** auf Hauptplatine oder im PCI-Slot zur Implementierung des SCSI-Protokolls
 - IDE ist einfacher zu implementieren und hat einen größeren Absatzmarkt (Geräte mit IDE/ATA Interface sind **preiswerter** als mit SCSI)
 - unterschiedliche Steckervarianten
- ⇒ SCSI ist führendes Bussystem im High-End Bereich (z.B. für Server oder Hochleistungs-Workstations)

4.1 CD-ROM

- Entwicklung von Philips und Sony (1984, *Yellow Book*)
- Aufbau einer CD (Audio +CD-ROM):
 - 5 ¼ Zoll Plastikscheibe mit Substrat, überzogen mit reflektierender Silberschicht und transparenter Schutzschicht
 - Substrat erhält durch Herstellungsprozess in der Spur eine Serie von Löchern (*Pit*) und Plateaus (*Land*)
 - **spiralförmige Spur**, innen beginnend, 22200 Windungen, 1.6 µm Breite, 5.68 km Länge
 - Abtastung mit Laserstrahl:
Land: gute Reflexion
Pit: schlechte Reflexion
 - Kodierung: jeder Wechsel
Land → *Pit* und *Pit* → *Land* repräsentiert eine „1“



4.1 CD-ROM (2)

- Organisation der Daten auf CD-ROM:
 - **Symbol**: jedes Byte wird mit 14 Bit kodiert
(Grund: *Pit* bzw. *Land* müssen eine Mindestlänge aufweisen)
 - **Frame** enthält 588 Bit
(24 Byte Nutzdaten, 8 Byte für Fehlerkorrektur, Rest für Synchronisation)
 - **Sektor** besteht aus 98 Frames
(16 Byte Präambel, 2048 Byte Nutzdaten, 288 zusätzliche Byte für Fehlerkorrektur)
⇒ für 2048 Byte Nutzdaten sind insgesamt 7023 Byte erforderlich!
- Abtastung mit linearer Geschwindigkeit von 1.2 m/s
- Transferrate:
 - 1x-Laufwerk: 75 Sektoren/s (153.6kBytes/s)
 - 52x-Laufwerk: 3000 Sektoren/s (7.98 MByte/s)
- Speicherkapazität beträgt ca. **650 MByte**